Probabilistic outlier detection in vibration spectra with small learning dataset

Aurélien Hazan, Michel Verleysen, Marie Cottrell, Jérôme Lacaille, Kurosh Madani

Probabilistic outlier detection in vibration spectra with small learning dataset

Aurélien Hazan[a], Michel Verleysen[b,c], Marie Cottrell[b], Jérôme Lacaille[d], Kurosh Madani[a]

[a] *LISSI - Université Paris-Est Créteil*
[b] *SAMM - Université Paris 1*
[c] *ICTEAM - Université Catholique de Louvain*
[d] *SNECMA - Division Intégration, Villaroche*

Abstract

The issue of detecting abnormal mechanical vibrations from spectra is addressed in this article, when little is known both on the mechanical behavior of the system, and on the characteristic patterns of potential faults.

With vibration measured from a bearing test rig and from an aircraft engine, we show that when only a small learning set is available, probabilistic approaches have several advantages, including modelling healthy vibrations, and thus ensuring fault detection.

To do so, we compare two powerful algorithms: the first one relies on the statistics of the maximum of log-periodograms. The second one computes the probability density function (pdf) of the wavelet transform of log-periodograms, and a likelihood index when new periodograms are presented. A by-product is the ability to generate random log-periodograms according with respect to the learning dataset.

Receiver Operator Characteristic (ROC) curves are built in several experimental settings, and show the superiority of our algorithms over state-of-the-art machine-learning-oriented fault detection methods; lastly we generate random samples of aircraft engine log-periodograms.

Keywords: fault detection, monitoring, vibration, rotating machine, bearing, aircraft engine, bayesian inference, periodogram, maximum, excess, wavelet, generalized pareto, probabilistic model

Email address: aurelien.hazan@u-pec.fr (Aurélien Hazan)

1. Introduction

We tackle the issue of health monitoring for rotating machines. Our application goal is the monitoring of vibrations in aircraft engines, but simpler test cases are also dealt with, for example bearing test rig. The following hypotheses are made:

- small learning set: about twenty short time-series are available, as a reflect of industrial constraints. More specifically, faulty data are scarce, if any.
- model-free: no specific mechanical model of the system, nor model of faults that might occur are used.
- constant target rotation speed: the rotating machines studied have a fixed speed. On short time intervals, the signal will be deemed to be stationary, so that periodograms are meaningful.
- nonparametric estimation: no specific functional form is assumed concerning periodograms, which will be decomposed in a wavelet basis.

In the spirit of many works in novelty detection [1] where information on faulty data is limited, our aim is first to come up with a nonparametric model of a healthy signal using a small learning set and secondly to compare a new signal to this model; the final goal is to detect unusual behaviors.

To build the model of healthy vibratory signals, we consider the log-periodograms of accelerometric signals and compare the discriminative power of two recently introduced algorithms [2, 3].

The first one models the density of an excess value of log-periodograms thanks to results from Extreme Value Theory (EVT, see [4]). The second one uses the wavelet transform of log-periodograms, which offers enough freedom in the perspective of function approximation. We stress the importance of the probabilistic description, be it in a Bayesian of frequentist framework, such that the periodogram models have an explicit form, that can be discussed and interpreted.

Results will be summarized using Receiver Operating characteristics [5, 3.4], which is a parametric plot of the False Positive Rate and the False Negative Rate with respect to a detection threshold t. FPR and FNR can be estimated if the data is annotated, i.e. if for each vibration record in the test dataset a label (fault/no fault) is available. Then for every threshold t,

we get a point (FPR,FNR). The ROC is widely used in radar, image and biomedical communities, but so far not as popular in vibration monitoring, as noticed for example in [6].

Section 2 links our work to related articles in various fields. Sections 3 and 4 present the main algorithms we use, whose results are summarized in Section 5. Section 6 concludes this article and discusses its perspectives.

2. Related work

Vibratory Health monitoring [7] involves mechanical science and signal processing. Signals may be studied in various domains: the time domain, the Fourier basis via STFT, the wavelet domain [8], or by other time-frequency distributions such as Wigner-Ville [9]. Over the years, it is also increasingly relying on machine learning and statistics [10, 11].

Condition monitoring of rotating machines often focusses on specific faults, such as rotor/stator contact [12], rotor unbalance, blade defects [13], bearing [14] and gearings defects [15]. However, unexpected problems can occur, with unknown fault patterns. Such concerns are germane to those developped in the area of novelty detection, where the importance of data not seen during the learning phase is stressed. Facing this problem, the best solution found by many authors is to build a model of normality, for example with neural networks such as Self-Organizing Maps [16, 17, 18]. This approach is sometimes termed *generative*, in contrast with a *discriminative* one [10].

Probabilistic approaches exist [1] to model normal behaviors. For example, a Bayesian approach to normality modelling in jet engine health monitoring has been developped [19, 20, 21]. The authors show that using Extreme Value Theory to model the maxima of order amplitudes increases the robustness of the detection procedure.

So far, these works address the case of a restricted number of shaft order amplitudes, and not the whole periodogram. In Sections 3 and 4 we discuss algorithms that belong to the probabilistic generative approach to novelty detection, in the case of vibrations monitoring in the spectral domain where the dimensionality of data is high.

3. Algorithm POT: peak-over-threshold statistics for log-periodograms

The aim of this algorithm is to make the most of EVT in novelty detection -as spearheaded by Tarassenko, Clifton and co-workers [11]- but in the high-dimensional context of vibratory log-periodograms. In [11], the pdf of the

maximum of a given statistics is considered instead of the pdf of the statistics itself (e.g. the scalar energy of the first order of the low-pressure shaft of an aircraft engine). The parameters of this pdf are learnt thanks to Bayesian inference, before running a statistical test.

Our case is different because we deal with large vectors (the log-periodograms) rather than scalar or low-dimensional vectors [22]. We propose to use excess-value statistics instead of maximum statistics: given a vector threshold, i.e. an upper limit for the spectra, we claim that monitoring all peaks that go beyond this threshold can solve the problem. This step stems from well-known fault detection algorithms in vibration monitoring [7, 4.2], where a mask is built using healthy vibration data. However, this procedure lacks a probabilistic translation so far.

To do so, EVT provides us with the necessary tools, since it models the probability of the excess value $P(X \mid X > t)$. Under mild conditions on the pdf of X, if t is large enough then $P(X \mid X > t)$ can be approximated by the Generalized Pareto distribution [4, 5.3.1]:

$$F(x) \quad = \quad 1 - \left(1 + \frac{\gamma x}{\sigma}\right)^{-\frac{1}{\gamma}} \tag{1}$$

where γ is the shape and σ the scale, and both need to be estimated from measurements.

The fault detection algorithm may then be written:

1. select a subset of the learning dataset, made of N log-periodograms of length F. For each frequency f we compute the max of the log-periodograms across the subset. A real vector $m = [m_1, \ldots, m_f, \ldots m_F]$ is obtained, the *mask*.
2. spot excesses over the mask in the rest of the learning dataset. Only excess values $Y = X_f - m_f \mid X > m_f$ are recorded, regardless of the frequency for which they occur. They consitute a sample of scalar excesses $\{Y_i\}_{i \leq I}$, and serve as inputs to the parameter estimation of the Generalized Pareto distribution.
3. set a detection threshold t according to standard probabilistic considerations and define a decision rule: any excess Y over the threshold t is considered as a fault.

For new uncategorized data, the last two steps of the procedure are repeated: excesses Y over the mask are first computed, then compared to t. Fig. 1 summarizes the algorithm.

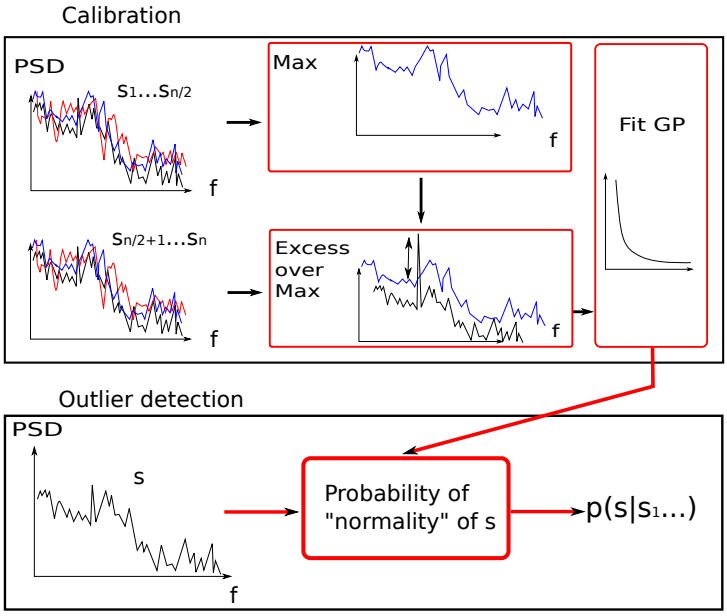


Figure 1: Peak-over-threshold detection algorithm.

Notice that we get round the difficulty of manipulating multivariate extremes, and at the same time take into account the frequency dependence of the spectra thanks to the mask. Parameter estimation and the decision rule are cast in a frequentist way, but could also follow Bayesian guidelines, as in Section 4.

4. Algorithm BW: Bayesian detection in a wavelet basis, random spectra generation

In this Section, we propose to build a probabilistic model of normality of log-periodograms in the wavelet domain as a means to detecting novelty as illustrated by Fig. 2. This is justified by the fact that such models were developped in statistical time series analysis and signal processing, for spectrum denoising purposes (Moulin [23], Percival and Walden [24, 10.6], Vidakovic [25, 9.3], Pensky et al. [26]), via wavelet thresholding or shrinking. The motivation of researchers in this area mainly concern the statistical properties of estimators (such as fixed or variable bandwith smoothing), which will not be discussed here. However we propose to take advantage of the model of normality that is provided by their analysis.

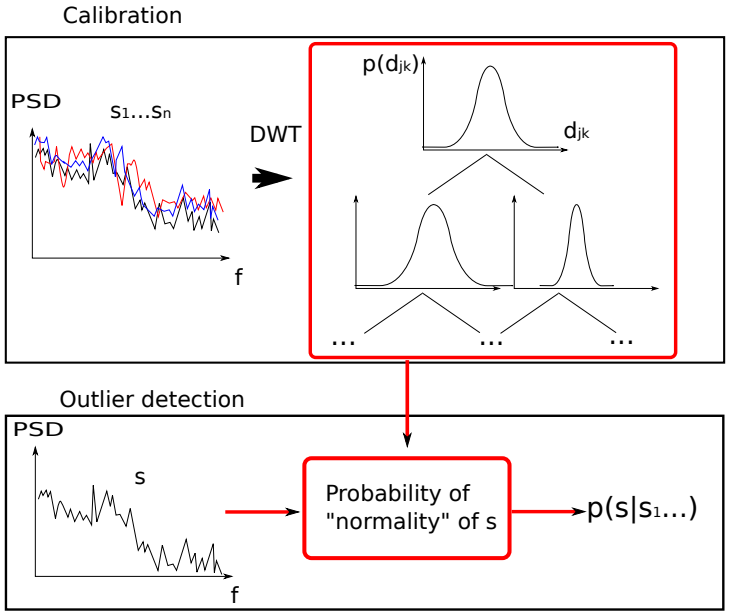


Figure 2: Bayesian detection in a wavelet basis (BW).

Before continuing, let us briefly discuss two objections that might be formulated:

- Why not working directly with the wavelet transform of the time-domain signals, without computing its periodogram ? The answer is that no simple probabilistic model of the coefficients would be available.
- Why not working directly with the probabilistic models of the log-periodograms ? Indeed, as will be seen in Section 4.1, such a probabilistic model is available, under stationarity assumptions. However, the noise distribution is not standard (see Section 4.2), which complicates subsequent computations.

Basic knowledge of the wavelet transform, and of its discrete implementation is assumed in this section. Theorical foundations, principles of fast computation, as well as practical illustrations may be found in [27].

4.1. Probabilistic model of the wavelet transform of a periodogram

What follows is standard material, available from [23]. We adopt the notations used in [26].

Let $I(\omega_j)$ be the periodogram at Fourier frequency $\omega_j = \frac{2\pi j}{T},\ j \in [0, T-1]$ associated with the vibration signal $X_0, \ldots, X_{T-1}$:

$$I(\omega_j) \quad = \quad \frac{1}{2\pi T} |\sum_{t=0}^{T-1} X_t e^{-i\omega_j t}|^2 \tag{2}$$

$I(\omega_j)$ is an estimator of the power spectrum density (PSD) which probability density function can be approximated under mild stationarity assumptions [28] as a function of the true PSD $f(\omega_l)$:

$$I(\omega_l) \quad \approx \quad \frac{1}{2} f(\omega_l) \chi_2^2 \tag{3}$$

where $l \in [2, T-2]$, χ_2^2 is a chi-square random variable with parameter 2. This term emerges as the sum of two squared Gaussians, one for the real part and one for the imaginary part of the discrete Fourier transform.

Taking the log of (3), a regression formula can be proposed :

$$z_l \quad = \quad \ln f(\omega_l) + \varepsilon_l,\ \forall l \in [2, T-2] \tag{4}$$

where $z_l = \ln I(\omega_l) + \gamma$, γ is Euler's constant, and ε_l are real-valued independent random variable with density μ. Equation (4) is called a regression formula because in practice $\ln f(\omega_l)$ is the unknown and must be estimated using z_l. It can be shown that:

$$\mu(x) \quad = \quad \gamma^* \exp(x - \gamma^* e^x) \tag{5}$$

$$E[\varepsilon_l] \quad = \quad 0 \tag{6}$$

$$V[\varepsilon_l] \quad = \quad \frac{\pi^2}{6} \tag{7}$$

where $\gamma^* = e^{-\gamma}$.

The discrete wavelet transform of eq. (4) gives:

$$d \quad = \quad \theta + \delta \tag{8}$$

where:

$$d \quad = \quad W[z_1, \ldots, z_T] \tag{9}$$

$$\theta \quad = \quad W[\ln f(\omega_1), \ldots, \ln f(\omega_T)] \tag{10}$$

$$\delta \quad = \quad W[\varepsilon_1, \ldots, \varepsilon_T] \tag{11}$$

and W is an orthogonal matrix given by the discrete wavelet transform. d, θ and δ may also be indexed by (j,k), where j is the scale and k the position. d_{jk} and θ_{jk} will thus be the scalar value at index (j,k) in vectors d and θ.

By Central Limit Theorem arguments, the density of coefficients of vector δ can be approximated by a normal law, except for small scales where a correction must be applied.

4.2. Bayesian inference of a log-periodogram, in the wavelet domain

Assuming the model of Section 4.1, what can be learnt from measurements on the distribution of the wavelet coefficients θ ?

Here we assume a Bayesian inference scheme, since it ensures that θ is modelled as a random variable rather than a fixed value. Prior for the wavelet coefficient θ_{jk} of the following form may be found in the litterature [26]:

$$\theta_{jk} \quad \rightsquigarrow \quad \pi_j \delta(0) + (1-\pi_j)\tau_j \xi(\tau_j \theta_{jk}) \tag{12}$$

where ξ is symmetric (such as a normal law $\mathcal{N}(0,1)$), and π_j, τ_j are hyperparameters. They can be learnt independently, taking advantage or theoretical arguments [26, 2.1]. In this work, we simply use a centered normal prior with variance obeying the following model [29]:

$$\sigma^2 \quad = \quad C 2^{-\alpha j} \tag{13}$$

where C and α are constants learnt from the data. The decrease of σ^2 as an inverse power of the scale j is consistent with theoretical results available under regularity hypotheses. More informally, if the function we approximate in the wavelet basis is smooth enough, then the dispersion of the wavelet coefficients decreases exponentially with j.

Let $\{d^{(1)}, \ldots, d^{(n)}\}$ be a set of n vector-valued wavelet-transformed log-periodograms, as defined in Equation (9). A posterior can be computed by the classical Bayes formula:

$$P\left(\theta_{jk} | d^{(1)}_{jk}, \ldots, d^{(n)}_{jk}\right) \quad \propto \quad l_{\mathbf{d}_{jk}}(\theta_{jk}) Pr(\theta_{jk}) \tag{14}$$

where $d^{(i)}_{jk}$ is the scalar indexed by (j,k) in the i-th vector-valued wavelet-transformed log-periodogram, $l_{\mathbf{d}_{jk}}(.)$ is the likelihood of d_{jk} under the normal noise model discussed in 4.1, and $Pr(.)$ stands for the probability of a given random variable.

Standard calculus shows that the posterior has the following form:

$$\forall (j,k), P\Big(\theta_{jk}|d^{(1)},\dots,d^{(n)}\Big) \;\propto\; \exp\Big(-\frac{1}{2\sigma_0^2}\Big[\theta_{jk}-\frac{\widehat{d}_{jk}}{1+\frac{\sigma_1}{\sigma_2}}\Big]^2\Big) \quad (15)$$

where $\widehat{d}_{jk}$ is the mean wavelet coefficient of the sample periodograms, and $\sigma_0, \sigma_1, \sigma_2$ are standard deviations whose formula are given in Appendix A.

σ_2 can be interpreted as the strength of the signal, and σ_1 as the strength of the noise. Hence $\frac{\sigma_2}{\sigma_1}$ can be thought of as the Signal to Noise Ratio (SNR). It can be seen that the higher the SNR, the closer the posterior will be to the empirical mean $\widehat{d}_{jk}$. Conversely, when the SNR is low, the empirical mean is a less trusted estimate. This is the desired behavior of the Bayesian estimator, designed to balance the empirical mean when little data is available.

4.3. Random generation of log-periodograms

Once the distribution of the posterior in Equation (15) is computed thanks to log-periodogram samples, one can sample from this distribution. Computing the inverse wavelet transform, we get a random log-periodogram sample. Examples will be given in Section 5.

Due to acquisition cost, such random samples can be of high interest to test detection algorithms. The classical bayesian fault detection procedure is highlighted in the following section.

4.4. Fault detection

So far we have chosen an estimation model (see Eq .(8)), proposed a prior and a posterior (see Eq. (15)). We can now compute the marginal likelihood (Bishop [30, Eq.(3.67-68)], Clifton et al. [20, Eq.(5)]), which quantifies the likelihood of a new sample, given the training set

$$d \mapsto p(d|d^{(1)},\dots,d^{(n)}) \;=\; \int p(d|\theta)p(\theta|d^{(1)},\dots,d^{(n)})d\theta \quad (16)$$

The integral may be approximated by Monte-Carlo sampling [31, 3.2]. If $p(d|d^{(1)},\dots,d^{(n)})$ is below a given threshold, a fault is suspected to occur.

Finally the detection algorithm reads:

1. Learn the model for wavelet coefficients:
   (a) Let $\{d^{(1)},\dots,d^{(n)}\}$ be a set of wavelet transforms of log-periodograms.

(b) For each (j,k), compute the mean and variance of $\theta_{jk}|d^{(1)},\ldots,d^{(n)}$ using Equation (15).

2. Detect:
   (a) Let d be the wavelet transform of a new log-periodogram, compute the marginal likelihood using Equation (16).
   (b) Set a detection threshold t. If $p(d|d^{(1)},\ldots,d^{(n)}) < t$, detect a fault.

Fig. 2 summarizes the algorithm.

5. Data and Results

5.1. Data: labelled IMS bearing dataset

The IMS bearing dataset [32] is a publicly available[1] set of vibration signals. Four bearings are installed on a shaft that rotates at a constant speed of 2000 rpm. Progressive degradations are recorded over a month from 8 accelerometers as the designed life time of the bearings is exceeded.

Log-periodograms with length $T = 8092$ are displayed by Fig. 3 at the beginning and at the end of the test, when a bearing is damaged.

Two datasets are built from the IMS recordings: one learning dataset, with 25 snapshots taken at the start of the recording session, while all bearings are healthy. Then, a test dataset is designed with 50 new recordings, 25 taken at the start of the test and 25 after n days of operation when light damage appear. The higher n, the easier the detection task because of the fast degradation of the bearing.

The test dataset is termed "labelled" because it comes with a label (fault/no fault) for each vibration. This feature is necessary to compute ROC curves, as in Section 5.2.

5.2. Fault detection with Algorithm POT

We first discuss the estimation of the parameters (σ,γ) of the probability of the excess value Y in Eq.(1). This estimation can be done with a frequentist [4, 5.3.2] of Bayesian [4, 11.5.3] point of view. Here we choose the frequentist approach, implemented in Matlab[2].

[1] `http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/`

[2] see the function gpfit in Statistics Toolbox.

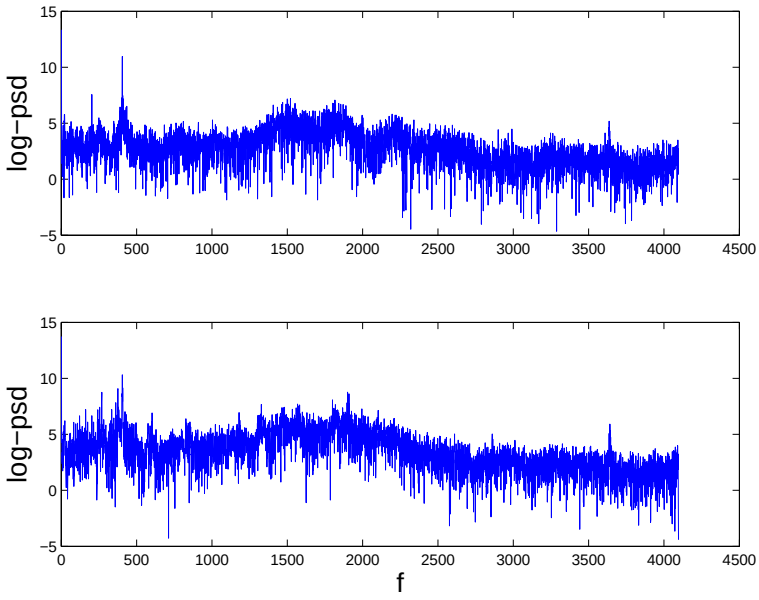


Figure 3: Log-periodogram of bearing vibrations (top) at the beginning of the test; (bottom) at the end of the test when a bearing is damaged.

Fig. 4 shows a good accordance between the histogram of excesses Y defined in Section 3 and the fitted pdf. It is an important result that guarantees the quality of further processing steps. The estimated parameters[3] are $\gamma = -0.05$, $\sigma = 0.72$.

Secondly, we plot the ROC curve of the detection algorithm in Fig. 5, defined here as the empirical false negative rate (FNR) as a function of the empirical false positive rate (FPR). Both FPR and FNR are functions of the detection threshold t defined in Sec. 3. There is a classical tradeoff between the two rates, in the sense that it is not possible to decrease arbitrarily the two rates simultaneously while moving t. Faulty data are recorded just $n = 2$ days after the beginning of the fatigue test, which explains why the ROC curve does not approach the origin very closely. Comparisons will be made with other algorithms in Sec. 5.3.

Lastly we examine the behavior of both FNR and FPR with respect to the threshold t, to make explicit the dependence and show how it should be

[3] negative γ distributions are referred to as the class of "extremal Weibull" distributions in the EVT literature.

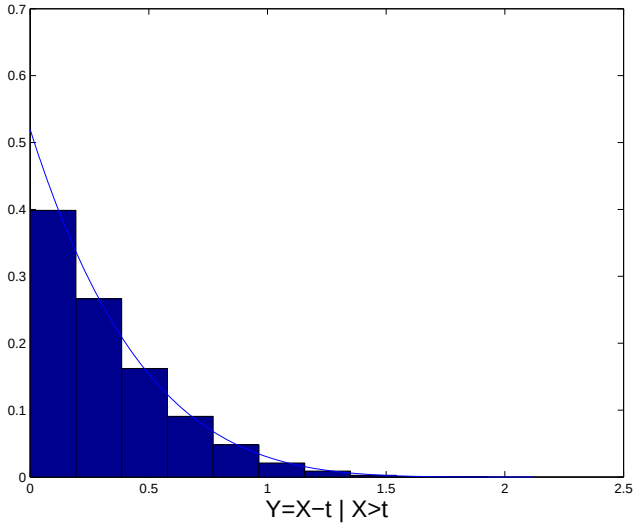


Figure 4: Histogram of excess values $Y = X - t|X > t$ and fitted Generalized Pareto pdf. Estimated parameters are: $\gamma = -0.05$, $\sigma = 0.72$

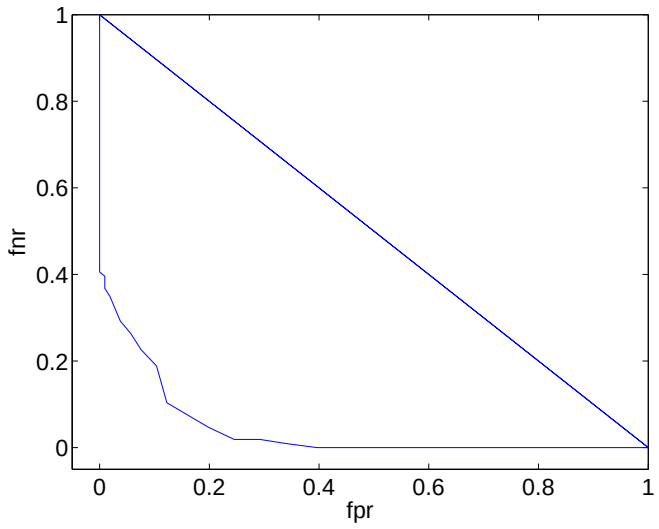


Figure 5: ROC curves for Algorithm POT (plain curve). Faulty test data recorded after $n = 2$ days of operation. The diagonal can be interpreted as the output of a detector which would randomly trigger an alarm, ignoring all the data.

chosen so as to minimize both error rates. We expect that as t increases, the FPR should descrease, while the FNR should rise. Fig. 6 reveals first that this is the observed behavior, and then that there is a single value of t close to 1.6 such that both error rates are low. It is impossible to minimize both at the same time, but a good compromise can be found. The obtained error rate -approximately 0.1- is high but is consistent with the fact that only 2 days have passed since the beginning of the fatigue test.

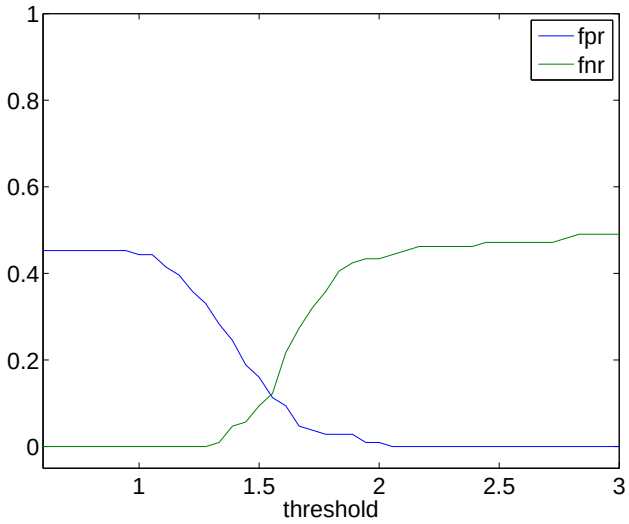


Figure 6: False Negative and False Positive Rates as functions of the threshold level t. Detection algorithm is POT.

5.3. Comparison of ROC curves

In this section the results of the POT and BW algorithms are compared with a state-of-the art discriminative novelty detection algorithm based on Kernel PCA [33], available online[4]. The principle of kPCA is to map data in a higher-dimensional embedding space, where it is easier to separate clusters linearly. In that space, a classical PCA can then be used. The detection algorithm first computes the kPCA of the learning dataset, then a residual measuring the distance between any new point and the principal components. When this residual passes a fixed threshold, a fault is detected.

[4] see `http://www.heikohoffmann.de/kpca.html`

Two parameters must be tuned for kPCA: the number of eigenvalues for the projection (here set to $d = 50$, out of $T/2 = 4096$) and the width of the kernel, set to $w = 10^{-2}$.

ROC curves are plotted in Fig. 7, with test data corresponding to the early phase of the fatigue test, after just $n = 2$ days of operation. It is noted that POT is almost everywhere the best of the three algorithms because is stands closer to the axes, except for some values. Then comes BW, and finally kPCA.

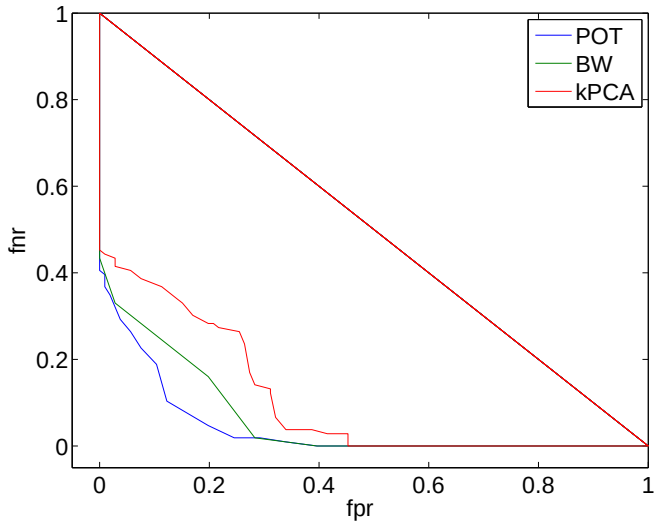


Figure 7: ROC curves for Algorithms POT, BW and kPCA. Faulty test data recorded after $n = 2$ days of operation. The diagonal can be interpreted as the output of a detector which would randomly trigger an alarm, ignoring all the data.

Then we focus on the POT algorithm and examine the influence of the dataset on the ROC curve. Fig. 8 compares ROC curves when faulty test data are recorded after $n = 2$ and $n = 7$ days of operation. The damage is more important after 7 days, thus the detection task should be easier, and the ROC curve closer to the axes [5, 3.4]. Remarkably, after 7 days the ROC curves is indistinguishable from the ideal shape of a perfect detector.

5.4. Random log-periodogram sampling with Snecma turbofan

So far, ROC curves have been obtained with the IMS bearing dataset. This is possible because the dataset is labelled. Unfortunately, in the context of aircraft engines, no labelled dataset is available to us so far. Consequently

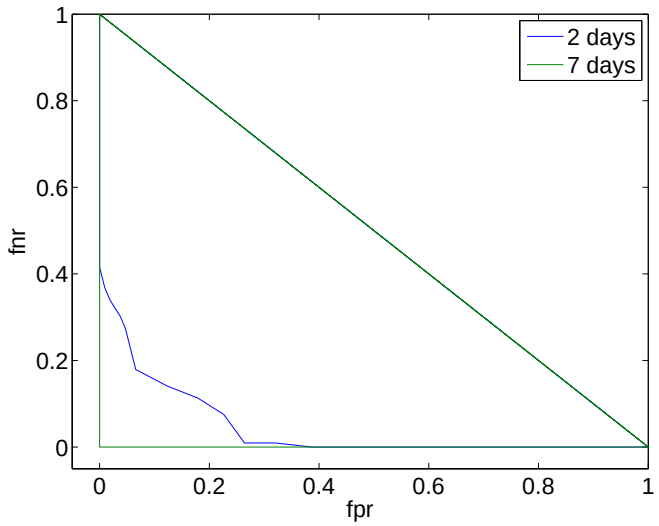


Figure 8: ROC curves for Algorithm POT, with faulty test data recorded after $n = 2$ and $n = 7$ days of operation.

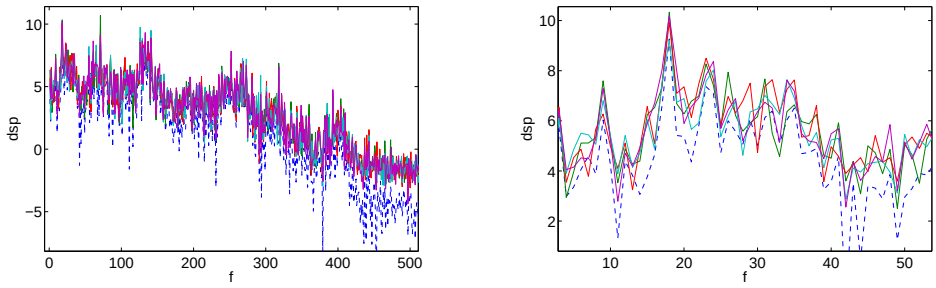
ROC curves can't be estimated. Nevertheless algorithm BW has an interesting feature that allows us to exploit Snecma data.

As mentionned in Section 4.3, a by-product of the inference of the pdf of the coefficients of the wavelet decomposition is the ability to generate random log-periodograms, conditionally on the learning dataset, and for a cost much inferior to that of performing a real test with an engine. This can be very useful to test any new algorithm in a signal processing workflow, particularly those requiring Monte-Carlo computations.

The recordings under study were provided by the Health Monitoring Department of Snecma[5] and correspond to a dual-shaft turbofan mounted on a testbench, that undergoes a continuous acceleration during several minutes. They include raw vibration outputs of two embedded accelerometers, sampled at 51kHz. Samples are collected while low-pressure shaft speed is at 2000rpm.

Randomly sampled log-periodogram are displayed in Fig. 9, which shows a good agreement between the learning set and random periodograms. In further works, we will show some precise applications of this random sampler.

[5] `http://www.snecma.fr`



(a) (b)

Figure 9: Randomly generated log-periodograms (dashed), conditionally on SNECMA aircraft engine learning dataset (plain lines), at various zoom levels.

6. Conclusion and perspectives

In this article the performances of three detection algorithms have been compared, in the context of vibratory condition monitoring in the spectral domain, with a small learning dataset, at constant regime and without mechanical model.

These constraints reflect those currently faced in the industry, for example in the field of large and complex rotating machines such as turbofans, where many defects can't be modelled nor anticipated, and where records of faults are very rare.

The two original methods belong to the class of generative probabilistic novelty detection algorithms. Peak-over-threshold (POT) is inspired both by classical practice in Condition Health Monitoring and by Extreme Value Theory. Bayesian detection in the Wavelet domain (BW) stems from research in signal processing and time-series analysis. They were compared to Kernel PCA (kPCA), which is a discriminative method.

POT, BW and kPCA were compared with respect to Receiver Operator Characteristic (ROC) curves, a well-established method used in many communities such as radar and biomedicine. The comparison was made with vibration measurements from a labelled bearing fatigue test, i.e. when we know whether a record corresponds to a fault or not. The importance of labels is stressed, because they are necessary to compute the False Positive and False Negative Rates in the ROC curves. Unfortunately, such labelling is very rare in aeronautic applications such as turbofan engines, since the default rate is below 10^{-7}.

However, in the case of IMS bearing dataset, it is shown that the generative model fits the spectral data well, and that both POT and BW novelty detection algorithms perform better than kPCA. Moreover we propose a way to randomly generate periodograms conditionally on a learning dataset. This feature is interesting to test other algorithms instead of recording new data, which is always expensive.

In future works we plan to :

- implement a Bayesian version of POT algorithm, and to evaluate the performance of the algorithms with related tools.
- compare the way we deal with multivariate data to other approaches.
- compare the performance of POT and BW with more novelty detection algorithms, and new data.
- generalize these techniques to variable regime.

Appendix A. Posterior

Here we give the expressions of the standard deviations that appear in the posterior eq.(15). σ_2 is the variance of the prior of the wavelet coefficient, which value is set according to eq. (13), which depends on j. σ_2 can be interpreted as the strength of the signal, and σ_1 as the strength of the noise.

We omit j subscripts for clarity:

$$\sigma_1 = \frac{1}{\sqrt{n}}\sqrt{\frac{\pi^2}{6}} \tag{A.1}$$

$$\frac{1}{\sigma_0} = \frac{1}{\sigma_1} + \frac{1}{\sigma_2} \tag{A.2}$$

where n is the number of samples, and σ_0 is the updated standard deviation, i.e. the standard deviation of the posterior.

References

[1] M. Markou, S. Singh, Novelty detection: a review. Part 1: statistical approaches, Signal Processing 83 (2003) 2481–2497.

[2] A. Hazan, M. Verleysen, M. Cottrell, J. Lacaille, Bayesian inference for outlier detection in vibration spectra with small learning dataset, in: Proceedings or Surveillance 6, Compiègne, France.

[3] A. Hazan, J. Lacaille, K. Madani, Novelty detection in vibration spectra using peak-over-threshold statistics, in: to be published in Proceedings of CFM-MPT'12, pp. –.

[4] J. Beirlant, Y. Goegebeur, Y. Teugels, J. Segers, Statistics of extremes. Theory and applications, Wiley, 2004.

[5] S. Kay, Fundamentals of statistical signal processing: detection theory, Prentice Hall, 1998.

[6] J. Nichols, S. Trickey, M. Seaver, S. Motley, Using roc curves to assess the efficacy of several detectors of damage-induced nonlinearities in a bolted composite structure, Mechanical Systems and Signal Processing 22 (2008) 1610 – 1622.

[7] R. Randall, Vibration-based Condition Monitoring, Wiley, 2011.

[8] Z. K. Peng, F. L. Chu, Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography, Mechanical Systems and Signal Processing 18 (2004) 199 – 221.

[9] J. Antoni, F. Bonnardot, A. Raad, M. E. Badaoui, Cyclostationary modelling of rotating machine vibration signals, Mechanical Systems and Signal Processing 18 (2004) 1285 – 1314.

[10] B. Ng, Survey of anomaly detection methods, Technical Report, Lawrence Livermore National Laboratory, 2006.

[11] L. Tarassenko, D. Clifton, P. Bannister, S. King, D. King, Novelty detection, in: K. Worden (Ed.), Encyclopaedia of Structural Health Monitoring, Wiley, 2009.

[12] Z. K. Peng, F. L. Chu, P. W. Tse, Detection of the rubbing-caused impacts for rotor-stator fault diagnosis using reassigned scalogram, Mechanical Systems and Signal Processing 19 (2005) 391 – 409.

[13] V. Kharyton, Fault detection of blades in blades of an aviation engines in operation, Ph.D. thesis, Ecole Centrale de Lyon, 2009.

[14] R. Orsagh, J. Sheldon, C. Klenke, Prognostics/diagnostics for gas turbine engine bearings, in: Proceedings of IEEE Aerospace Conference, pp. –.

[15] W. Wang, F. Ismail, M. Golnaraghi, Assessment of gear damage monitoring techniques using vibration measurements, Mechanical Systems and Signal Processing 15 (2001) 905 – 922.

[16] A. Ypma, R. Duin, Novelty detection using self-organizing maps, in: N. Kasabov, R. Kozma, K. Ko, R. O'Shea, G. Goghill, T. Gedeon (Eds.), Progress in Connectionist-Based Information Systems, Vol.II, Springer Verlag, 1997, pp. 1322–1325.

[17] L. Tarassenko, A. Nairac, N. Townsend, P. Cowley, Novelty detection in jet engines, in: Colloqium on Condition Monitoring, Imagery, External Structures and Health, IEEE, 1999, pp. 41–45.

[18] M. Markou, S. Singh, Novelty detection: a review. part 2: neural network based approaches, Signal Processing 83 (2003) 2499–2521.

[19] D. Clifton, Condition Monitoring of Gas-Turbine Engines, Technical Report, Department of Engineering Science, University of Oxford, 2006.

[20] D. Clifton, N. McGrogan, L. Tarassenko, K. S., P. Anuzis, K. D., Bayesian extreme value statistics for novelty detection in gas-turbine engines, in: Proc. of IEEE Aerospace, Montana, USA, pp. 1–11.

[21] D. Clifton, L. Tarassenko, Novelty detection in jet engine vibration spectra, IET Condition Monitoring (2009) –.

[22] D. Clifton, S. Hugueny, L. Tarassenko, Novelty detection with multivariate extreme value statistics, Journal of Signal Processing Systems 65 (2010) 371–389.

[23] P. Moulin, Wavelet thresholding techniques for power spectrum estimation, IEEE Transactions on Signal Processing 42 (1994) 3126–3136.

[24] D. Percival, A. Walden, Wavelets methods for Time Series analysis, CUP, 2000.

[25] B. Vidakovic, Statistical Modeling by Wavelets, Wiley, 1999.

[26] M. Pensky, B. Vidakovic, D. De Canditiis, Bayesian decision theoretic scale-adaptive estimation of a log-spectral density, Statistica Sinica 17 (2007) 635–666.

[27] S. Mallat, A wavelet tour of signal processing, 3rd ed., Elsevier, 2009.

[28] D. Brillinger, Time Series Data Analysis and Theory, Holden Day, San Franciso, 1981.

[29] F. Abramovich, T. Sapatinas, B. Silverman, Wavelet thresholding via a bayesian approach, J. R. Statisc. Soc.B 60 (1998) 725–749.

[30] C. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[31] C. Robert, G. Casella, Introducing Monte Carlo methods with R, Springer, 2010.

[32] H. Qiu, J. Lee, J. Lin, Wavelet filter-based weak signature detection method and its application on roller bearing prognostics, Journal of Sound and Vibration 289 (2006) 1066–1090.

[33] H. Hoffman, Kernel pca for novelty detection, Pattern Recognition 40 (2007) 863–874.